Request Request Request Request Request Request Request Request

PCI-SIG ENGINEERING CHANGE REQUEST

TITLE:	PCIe CEM Thermal Reporting
DATE:	November 7, 2016
AFFECTED DOCUMENT:	PCI Local Bus Specification Revision 3.0 PCI Express CEM Specification 3.0
SPONSOR:	Dell, Hewlett-Packard, Lenovo

Part I

1. Summary of the Functional Changes

This ECN specifies changes to the PCI Local Bus Specification Revision 3.0 and the PCI Express CEM Specification 3.0. Changes to the PCI Local Bus Specification cover a new VPD encoding and a 32-bit field. Changes to the PCI Express CEM Specification cover a series of graphs used to classify air flow impedance and thermal properties under varying conditions as well as the test figure and process to create these graphs for a given adapter.

Adapter types supported by this include all single-wide and double-wide PCIe adapters without integrated air movers, including full-high as well as low profile adapters (with a full high bulkhead plate). Adapters with an integrated air mover were not addressed due to the added complication of their integrated air mover in the overall platform's potential cooling redundancy.

2. Benefits as a Result of the Changes

These changes enable a common method for adapters to be thermally assessed and compared. This allows platforms using this information to determine if the adapter can be supported within the platform's cooling capabilities. Making more full use of the information these changes provide, platforms may also opt to dynamically adjust cooling mechanisms, e.g., cooling fan speed, to ensure adapters receive appropriate cooling at optimal power consumption. This will reduce the probability of thermal-related damage to adapters and platforms while minimizing power consumption.

3. Assessment of the Impact

These changes do not impact adapters that do not support the thermal reporting fields.

4. Analysis of the Hardware Implications

To support thermal reporting, adapter providers will need to construct a test fixture and perform testing per the methods specified. Adapter providers will need to communicate the results through the reporting field designated within VPD.

5. Analysis of the Software Implications

To support thermal reporting, software, e.g., firmware, will need to read and comprehend the new VPD encoding and corresponding field.

6. Analysis of the C&I Test Implications

No changes are required since these curves do not impact current interoperability testing.

Part II

Detailed Description of the change

"Modify PCI Local Bus Specification Revision 3.0 as follows:"

I. Vital Product Data

I.3.1.1 Read-Only Fields

"Add a two-letter keyword"

TR	Thermal Reporting	This optional keyword provides a standard interface for reporting four thermal reporting fields: AFI Level, MaxTherm, DTherm, and MaxAmbient. This keyword value is encoded as a 4-byte binary value in little endian order (byte 0 contains bits 7:0). This values contains the four fields as follows: AFI Level bits [3:0], MaxTherm bits [7:4], DTherm bits [11:8], MaxAmbient bits [19:12], and placed in bits 19:0. Bits 31:20 must be set to 000h. Field description is provided within the *PCI Express Card Electromechanical Specification 3.0*. Note that due to the character nature of the VPD encoding mechanism, this binary value is permitted to start on any byte boundary within the VPD.

"Add a new section 11 to PCI Express Card Electromechanical Specification Version 3.0 as follows:"

11. Adapter Thermal Reporting

Adapter components are influenced by the temperature and airflow velocity and volume present across these components. To enable enclosure management to comprehend whether a given adapter is able to be adequately cooled based on its dynamic operating conditions, an adapter is permitted to communicate its operating requirements over a range of environmental conditions through PCI Vital Product Data (VPD). Software reads these values to determine whether an adapter is able to be supported based on the adapter's thermal characteristics as well as determine how to optimize cooling mechanisms to ensure correct operation without expending excessive power.

Adapter Thermal Reporting uses three graph sets that each define up to 15 operating curves. Each curve corresponds to one numeric value communicated by the corresponding VPD field. Values associated with unspecified curves are treated as Reserved.

11.1. Airflow Impedance (AFI) Level

The intent of Airflow Impedance Level information is to identify adapters which, if in the same airflow path as other platform features to be cooled, may contribute to a total airflow impedance that impacts the cooling for the platform as a whole. Also, the installation of higher impedance adapters, especially if multiple such adapters are installed, in a platform may exceed the

capabilities for a given platform fan's performance curve. Through testing, platform developers determine what adapter AFI Levels the platform supports.

Impedance is categorized as one of N levels communicated by the Air Flow Impedance (AFI) Level field. Each level corresponds to one of the curves illustrated in Figure 11-1. The level is assigned based on the highest AFI level number which is still below the measured adapter AFI throughout the range of air flows. The AFI Level field is set to a value of 1 through 9. All other values are currently Reserved.

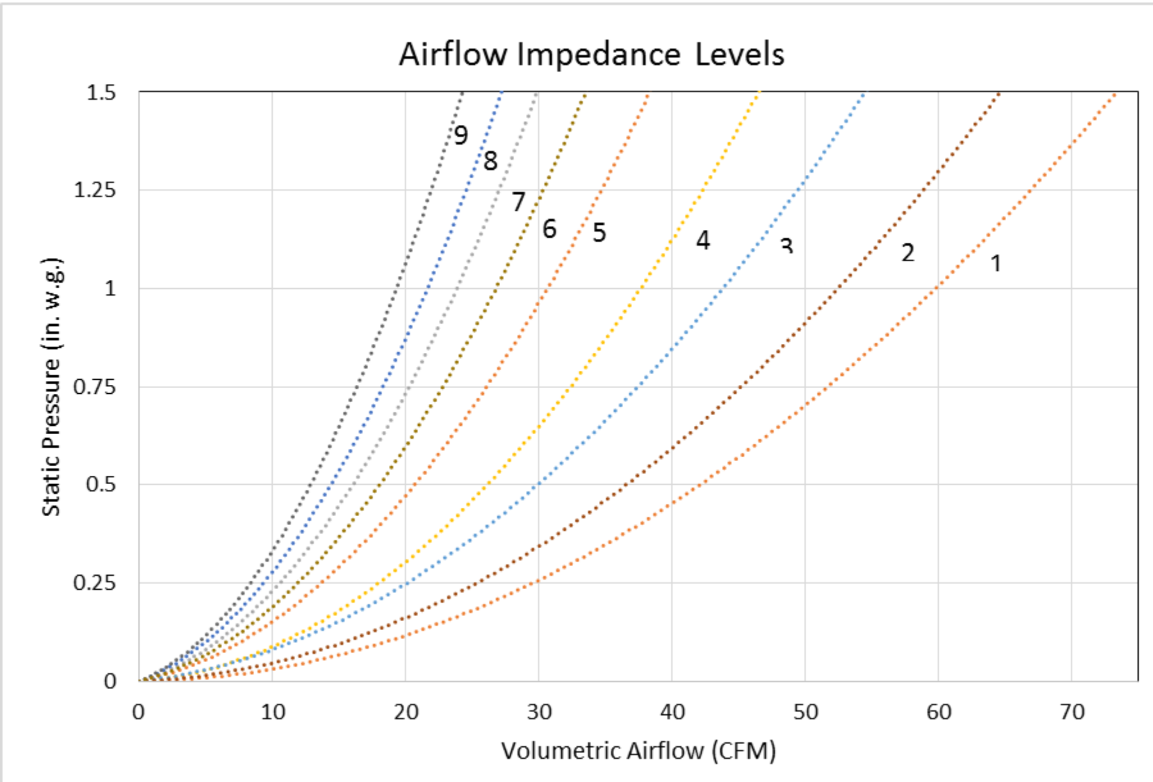

Figure 11-1: AFI Levels

An adapter is to be rated to it's appropriate AFI Level by comparing its measured impedance, in 0.25 inch water increments through the static pressure range of 0.25 to 1.25 in.w.g., to the AFI Level equations listed in Table 11-1 below. The lowest AFI Level which always has higher volumetric airflow for a given static pressure relative to the adapter card's measured impedance shall be listed as the adapter card's AFI Level.

The airflow impedance testing for an adapter beings at 0.25 in.w.g. to lessen the potential for error at low flowrates.

Table 11-1: AFI Level Equations.

AFI Level	Equation (X=volumetric airflow [CFM]; Y=static pressure [in.w.g.])
1	Y = 0.0003 * X^2 + 0.0003 * X
2	Y = 0.0003 * X^2 + 0.0012 * X

Request Request Request Request Request Request Request Request

3	Y = 0.0004 * X^2 + 0.0035 * X
4	Y = 0.0006 * X^2 + 0.0023 * X
5	Y = 0.0008 * X^2 + 0.0066 * X
6	Y = 0.0011 * X^2 + 0.078 * X
7	Y = 0.0014 * X^2 + 0.0091 * X
8	Y = 0.0016 * X^2 + 0.0117 * X
9	Y = 0.0020 * X^2 + 0.078 * X

11.2. Maximum Thermal (MaxTherm) Level

The intent of the Maximum Thermal (MaxTherm) Level information is to define the minimum airflow required at a given air temperature for which an adapter card, when stressed to its TDP level limit, will operate within its component's reliability limits and without degraded adapter performance.

Once an adapter's thermal performance profile is established (see Appendix C), the Maximum Thermal (MaxTherm) Level is determined. MaxTherm is the lowest curve in Figure 11-2 which is entirely above the adapter's quantified thermal performance curve throughout the range of approach ambient temperatures – also referred to as local ambient temperatures – shown on the X-axis ranging from 25°C up to 65°C.

Note that the curves shown in Figures 11-2 and 11-3 are for reference only. Compare measured readings against the equations defined in Table 11-2. Note that Thermal Levels 2, 3, and 4 have a low limit for their approach air speed of 100 LFM, meaning their Thermal Levels do not require an adapter to support less than 100 LFM.

It is suggested, but not required, that adapter card developers strive to offer solutions that can operate within the shaded recommended design space envelope shown in Figure 11-2 to maximize the number of systems that can support such conditions.

The MaxTherm field is set to a value of 1, 2, 3, 4, 5, 6, 7, or 8. All other values are currently Reserved.

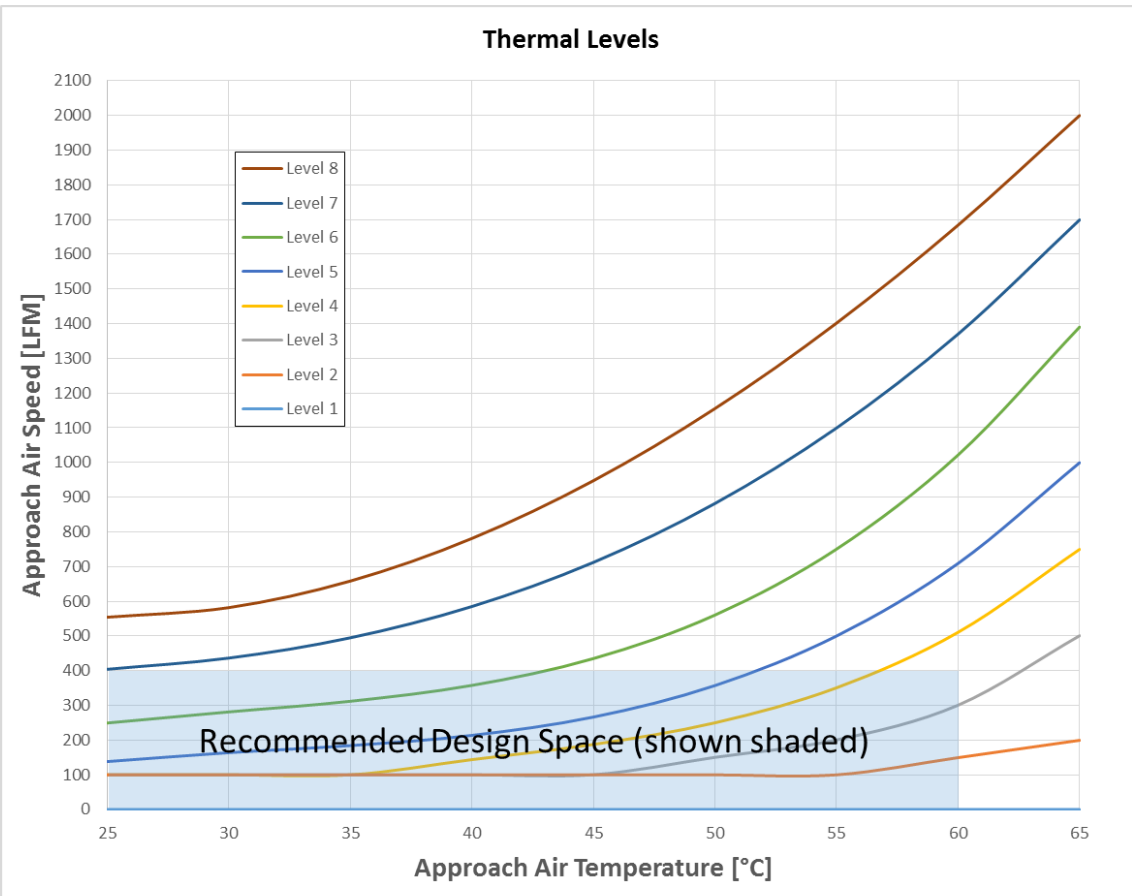

Figure 11-2: MaxTherm and DTherm Levels

No lower air speeds shall be presumed for approach air temperatures below 25°C.

For example, Figure 11-3 illustrates the curve of a hypothetical adapter that has been tested. Given this profile, the MaxTherm Level for the unit under test (UUT) adapter card would be set to 6 if it wished to support the entire temperature range up through 65°C since it exceeds MaxTherm Level 5 above an approach air temperature of 52°C.

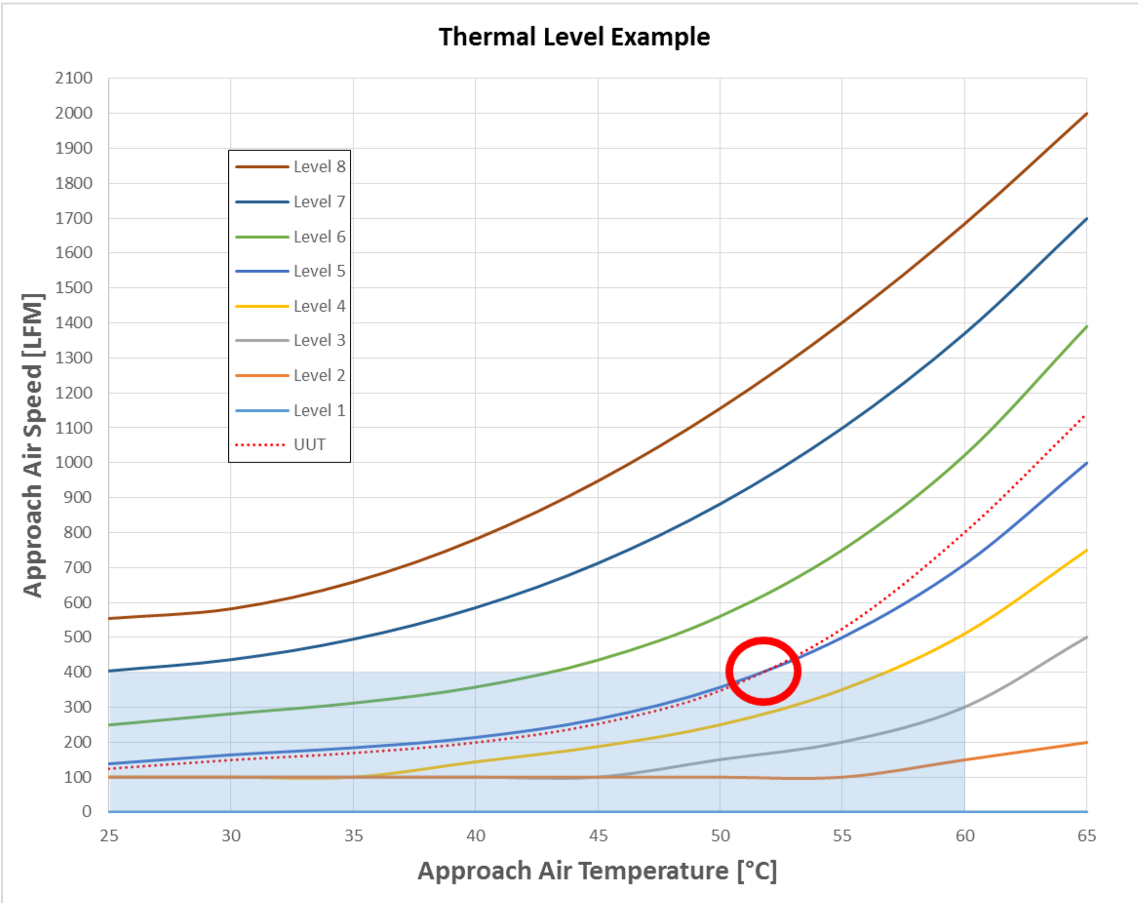

Figure 11-3: Example Adapter Thermal Profile

Table 11-2: Thermal Level Equations.

Thermal Level	Equation (X=approach air temperature [Celsius]; Y= approach airflow speed [LFM])
1	Y = 0
2*	Y = -0.00093 * X^3 + 0.16705 * X^2 – 250.92807
3*	Y = 0.06667 * X^3 – 10.0000 * X^2 + 508.33333 * X – 8600.00000
4*	Y = 0.02700 * X^3 – 3.27605 * X^2 + 140.74670 * X – 1972.30735
5	Y = 0.01915 * X^3 – 1.82974 * X^2 + 62.21147 * X – 572.75974
6	Y = 0.02201 * X^3 – 2.00909 * X^2 + 66.82862 * X – 509.54545
7	Y = 0.00717 * X^3 – 0.11926 * X^2 – 3.31674 * X + 449.64286
8	Y = –0.00340 * X^3 + 1.28248 * X^2 – 57.26411 * X + 1237.37229

Note: * Approach air speed values for indicated Thermal Levels shall not be less than 100 LFM.

11.3. Degraded Thermal (DTherm) Level

The intent of the Degraded Thermal (DTherm) Level information is to determine the minimum airflow required at a given air temperature for which an adapter card, when provided the same stress application as it was for MaxTherm, will operate within its component's reliability limits but at a degraded adapter performance level. Typically this is accomplished by the adapter's self-initiated thermal protection schemes, such as throttling. This is potentially of interest for platforms unable to provide sufficient cooling, such as due to the loss of a cooling fan which reduces the platform's cooling capacity. The reduced cooling capacity in such a case may be below the adapter's MaxTherm Level. For adapter's that are capable of lowering performance to thermally protect themselves, though their performance may be reduced they still remain viable plus do not risk causing the platform to shut down.

If an adapter is capable of operating at a reduced performance level to protect itself thermally and this translates into a reduced Adapter Thermal Profile then that established for MaxTherm, then DTherm is not equal to MaxTherm. DTherm is calculated using the same curves and process as MaxTherm, however, the adapter's profile is created using the adapter's most reduced performance operating level at which it is still on and providing useful work that is not less than 10% of its maximum sustained performance noted in MaxTherm testing. The DTherm field is set to a value of 1, 2, 3, 4, 5, 6, 7 or 8. All other values are currently Reserved.

11.4. MaxAmbient

Some adapters may not be thermally viable up to an approach air temperature of 65°C, regardless of the air speed involved. Or, even if the adapter could be thermally viable at high temperature extremes, the airflow requirements to do so may require it to operate at a higher thermal challenge level, potentially reducing the number of systems able to supply the needed airflow. In either circumstance, the adapter may opt to have its upper approach air temperature limit be less than 65°C. This approach air temperature upper threshold is referred to as MaxAmbient. MaxAmbient must be a value between 50 and 65 inclusive. It represents a temperature in degrees Celsius. Its value shall be entered as a whole number. It applies to the full operating potential MaxTherm condition.

For the example of establishing a MaxTherm Level for an adapter, as previously illustrated in Figure 11-3, if it was desired by the adapter's developer to be less than a MaxTherm Level of 6, then a MaxAmbient value of 52 (representing 52°C) could have been used by the adapter's developer instead of 65. In such a case, declaring a MaxTherm level of 5 with a MaxAmbient of 52 would have been acceptable. If MaxAmbient is unsupported, then these bits must be set to 0000 0000b.

"Add a new Appendix C to the PCI Express Card Electromechanical Specification Version 3.0 as follows:"

C. Thermal Data Collection and Test Procedure

This appendix describes the thermal test set up and procedure to evaluate a PCIe adapter's thermal performance. Thermal performance is communicated via four VPD fields—AFI Level, MaxTherm Level, DTherm Level, and MaxAmbient. Details regarding the dimensions and construction of the fixture are provided in the associated CAD files for the thermal challenge and airflow impedance tester.

The test fixture is intended to be attached to an AMCA 210-99/ASHRAE 51-1999 compliant airflow chamber, which can quantify both static pressure as well as volumetric airflow. The tester is attached to the airflow chamber such that air blows toward the PCIe adapters and exits the tester at the adapter's bulkhead plate.

For MaxTherm and DTherm level testing, the airflow impedance plates (shown as orange in color in Figure C-1) must be removed. For AFI Level testing, where the desire is to understand the relative airflow impedance of the adapter card, the use of the airflow impedance plates (two when testing single-wide adapters; one when testing double-wide adapters) is required.

The PCIe adapter to be tested is installed in the test fixture, which is positioned above a working IO board's PCI x16 connector. Cables, as needed to fully exercise the adapter, are attached to the adapter to be tested. The test fixture's lid is then attached.

The adapter is operated to its rated TDP level using an exercise procedure that the adapter's manufacturer must define in sufficient detail - including full descriptions of all software, firmware, and hardware needed - such that others could recreate the same results if provided the same adapter card. This information, including special software scripts if needed, is to be made readily available to those wishing to similarly test the adapter. Availability of this information is to align with the adapter's production availability.

See the associated CAD file information for details of the tester's construction and size. Figure C-1 shows an exploded view of the tester, with the lid, its mounting screws, the AFI blocks (shown in orange for this document's illustration purposes only), and adapter card under test shown suspended above the tester housing. Again, during Thermal Level testing for MaxTherm and DTherm data, the AFI blocks (shown in orange) MUST be removed from the tester. The tester is connected to a flow bench which provides quantifiable volumetric airflow and static pressure.

Request Request Request Request Request Request Request Request

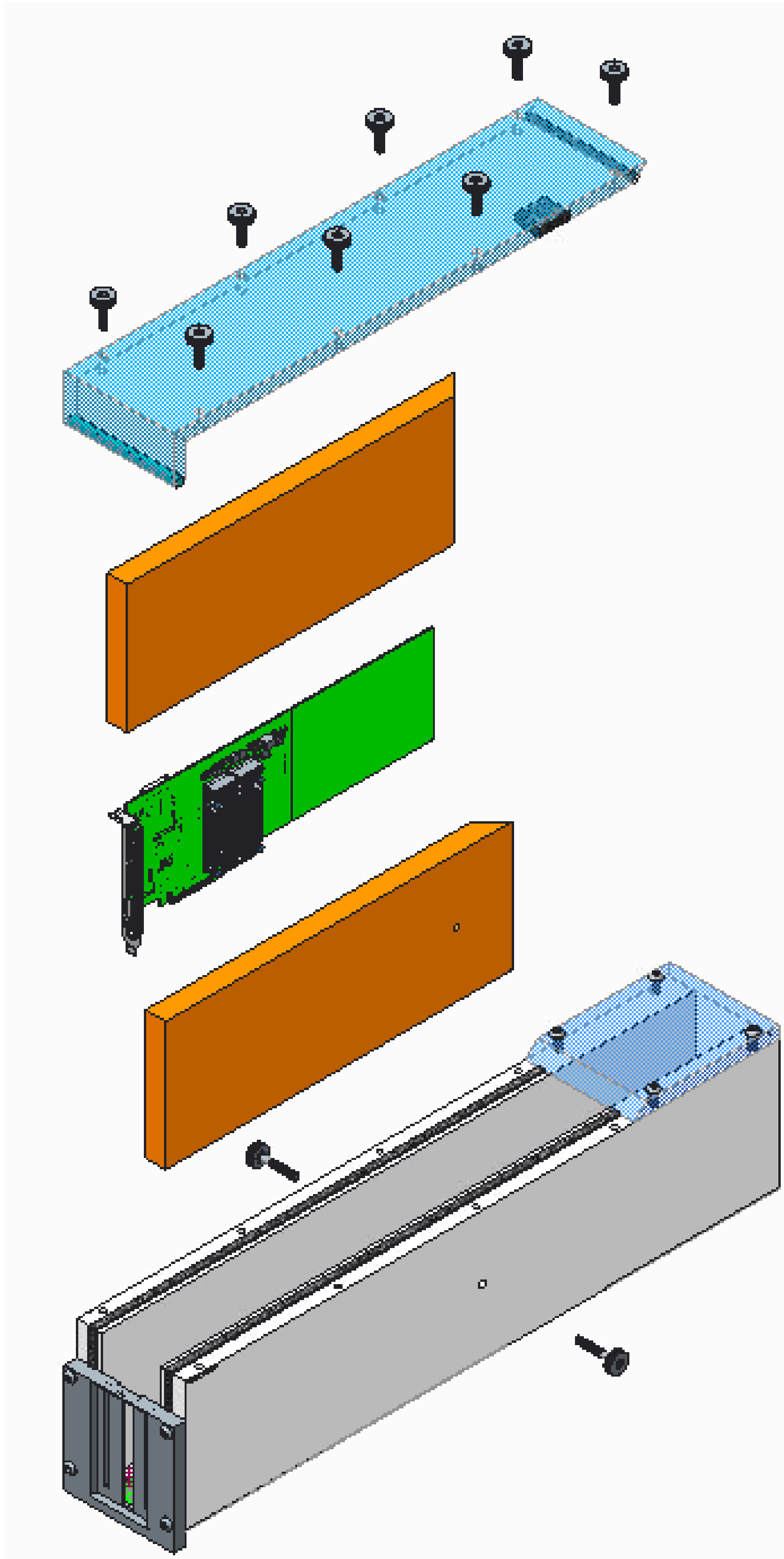

Figure C-1: Tester isometric exploded view, lid and impedance blocks removed

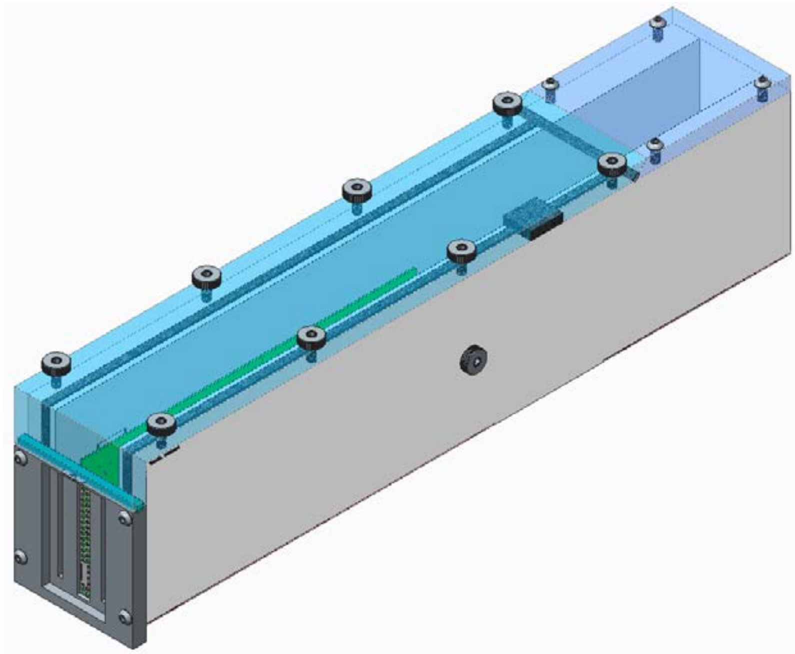

Figure C-2: Tester isometric view in MaxTherm and DTherm testing configuration

(Both AFI blocks removed)

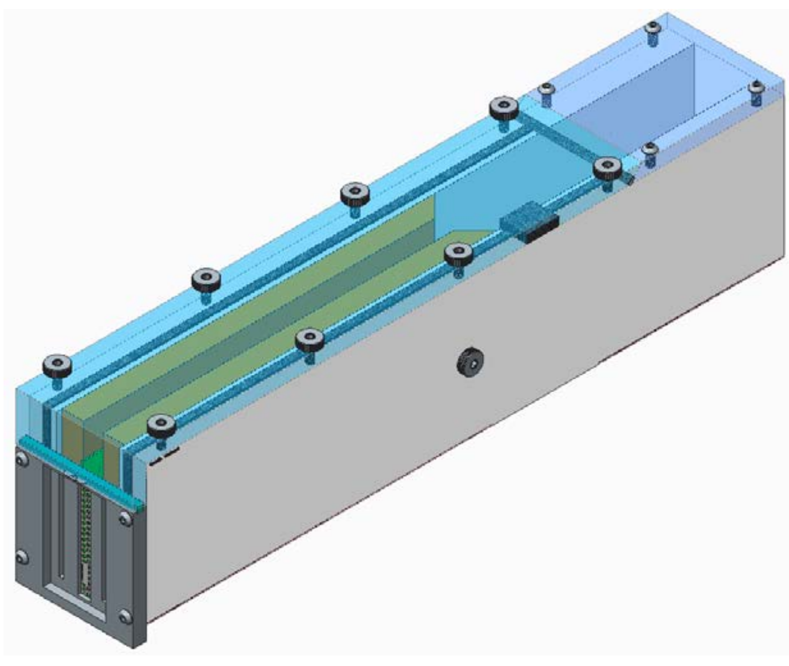

Figure C-3: Tester isometric view in single-wide adapter AFI testing configuration

(Both AFI blocks installed)

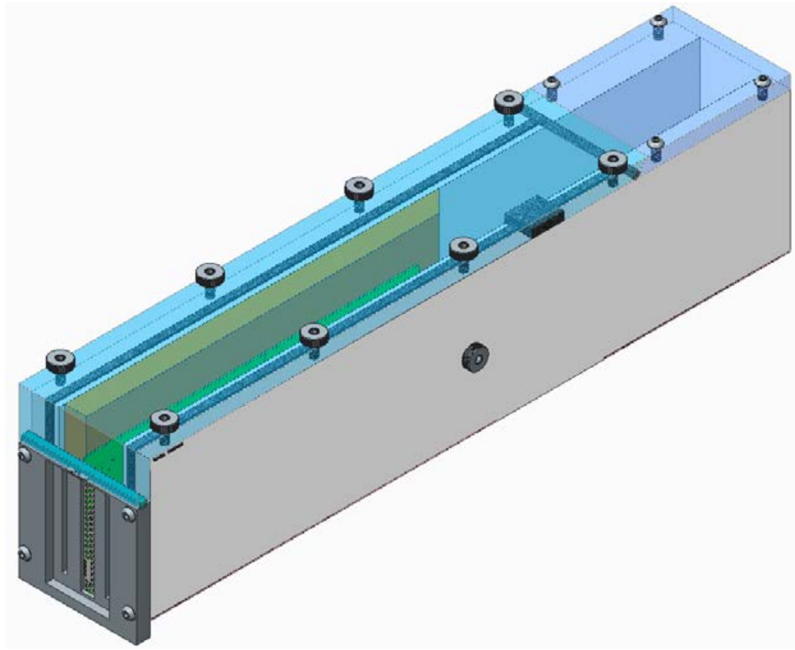

Figure C-4: Tester isometric view in double-wide adapter AFI testing configuration

(One AFI block installed)

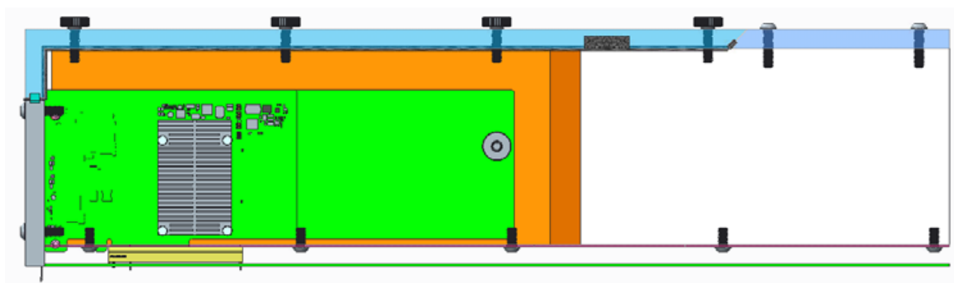

Figure C-5: Side view in double-side adapter AFI testing configuration

(Side wall removed for viewing illustration only)

A system board with one x16 PCI slot is needed for the testing. For MaxTherm and DTherm testing, the quantified volumetric airflow is divided by the cross-sectional area of the duct to establish the average velocity of air entering the test fixture: velocity (FT / min) = volume flow (ft^3 / min) / tester's inner area (ft^2) perpendicular to airflow. One hot-wire anemometer should be placed at the center of tester's air channel, within 50 mm downstream of the tester's air entrance opening. The anemometer data should be collected for reference only. At least one thermocouple sensor is required to be placed at this same location to collect inlet air temperature data, unless also using the anemometer's local ambient temperature reading. Figure C-6 shows an example test setup connected to a flow bench.

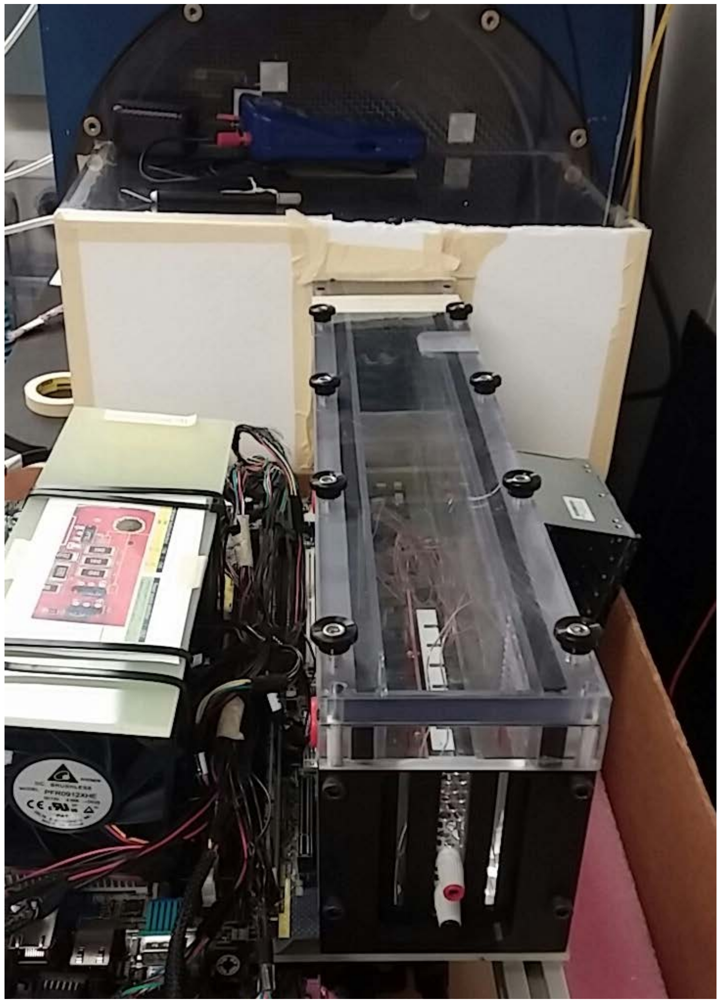

Figure C-6: Actual Thermal Level Test Fixture Set-up

The target output is the air speed (LFM, or FT / min) required to cool the adapter under test at projected (or actual) air temperatures while the adapter card is being stressed to its TDP level. The adapter's component's minimum temperature margin at various air speeds are collected, while also noting the air temperature entering the test fixture. The adapter's minimum temperature margin is added to the tester's air inlet temperature to arrive at a projection as to what would be the zero temperature margin ambient for given air speeds. With the exception of adapters meeting the requirements of Thermal Level 1, no less than six data points, regularly spaced from a project approach ambient of 25°C to at least 50°C (65°C preferred), are to be collected. These projected approach ambients and approach air speeds are then compared against the Thermal Levels shown in Figure 11-2.

For example, a hypothetical fabric adapter card was profiled for its thermal levels using an airflow chamber, such as the one shown in Figure C-7, at various air speeds.

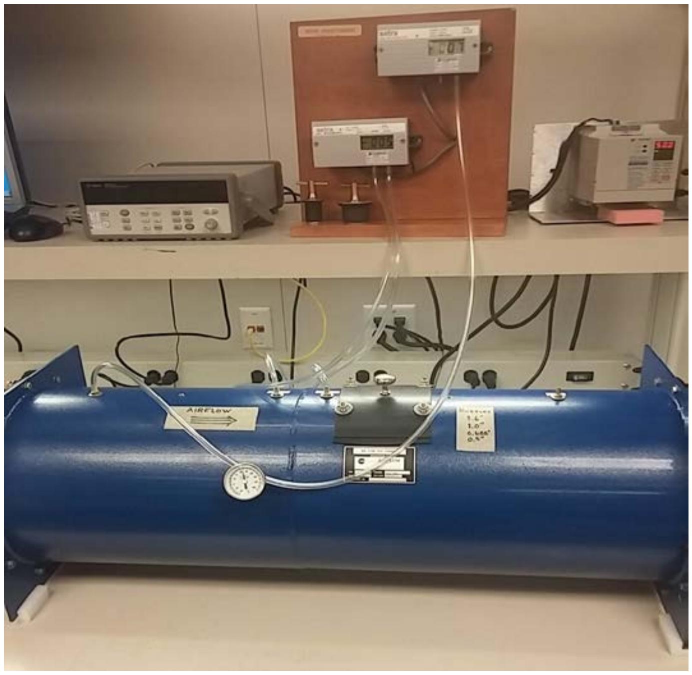

Figure C-7: An Example of an Airflow Chamber

Several measurement points are made to enable a good curve fit across the range of interest. The components of least temperature margin are monitored using temperature sensors that are part of the adapter's design, and added thermocouples, as needed. An optical transceiver plug, technically not a part of the adapter card, but being integral for its intended usage, is included in this adapter's characterization. The results for this hypothetical adapter when running at its full potential may look like the data in Table C-1.

Table C-1: Hypothetical Fabric Adapter Card Thermal Level Measurements

Hypothetical Fabric Adapter Card					
Test Run	**Approach Ambient Temperature (°C)**	**Measured CFM**	**VR Temperature (thermocouple reading: 100C max allowed) (°C)**	**QSFP Temperature (adapter sensor reading: 70C max allowed) (°C)**	**ASIC Temperature (adapter sensor reading: 105C max allowed) (°C)**
1	26	49	39.3	31	54.5
2	26	41	40	34	55
3	26	32	41	38	56
4	26	27	42	41	58
5	25	23	44	43	60
6	25	18.5	47	47	64
7	25	15	56	52	78
8	25	12	65	56	92
9	25	7.5	76	68	105

From these measurements the average air speed in LFM (linear feet per minute) can be calculated by dividing the volumetric airflow in CFM (cubic feet per minute) by the tester's inner channel area. The tester's inner dimensions are 4.724 inches in height by 2.337 inches in width. Table C-2 shows the results of the calculated LFM.

Table C-2: Hypothetical Fabric Adapter Card's Calculated Approach LFM

Hypothetical Fabric Adapter Card					
Test Run	**Approach Ambient Temperature (°C)**	**Measured CFM**	**Tester's inner channel height (IN)**	**Tester's inner channel width (IN)**	**Calculated LFM (CFM/tester's channel area)**
1	26	49	4.724	2.337	**639.1**
2	26	41	4.724	2.337	**534.8**
3	26	32	4.724	2.337	**417.4**
4	26	27	4.724	2.337	**352.2**
5	25	23	4.724	2.337	**300.0**
6	25	18.5	4.724	2.337	**241.3**
7	25	15	4.724	2.337	**195.7**
8	25	12	4.724	2.337	**156.5**
9	25	7.5	4.724	2.337	**97.8**

Request Request Request Request Request Request Request Request

The temperature margins at the different air speeds measured are calculated by subtracting the measured temperature for each component from its maximum allowed temperature for reliability and data integrity, sometimes referred to as the component's maximum continuous operating temperature, or MCOT. No additional temperature margin or safety factor should be used in these calculations.

By adding the test run's actual measured approach ambient to the least temperature margin for the adapter's components, the result is the maximum approach ambient supportable for that air speed. Note in this example how under different air speeds, different components may have the least temperature margin for the adapter.

.

Following the above described steps, the approach temperature to plot against the calculated LFM for this hypothetical adapter card may be determined, as shown in Table C-3.

Table C-3: Hypothetical Fabric Adapter Card's Calculated Approach Air Temperature

Hypothetical Fabric Adapter Card					
Test Run	**Approach Ambient Temperature (°C)**	**VR margin to 100°C (°C)**	**QSFP margin to 70°C (°C)**	**ASIC margin to 105°C (°C)**	**Max Approach Temperature at Calculated LFM [Approach Ambient + Least Margin Component] (°C)**
1	26	50.7	**39**	50.5	**65**
2	26	50	**36**	50	**62**
3	26	49	**32**	49	**58**
4	26	48	**29**	47	**55**
5	25	46	**27**	45	**52**
6	25	43	**23**	41	**48**
7	25	34	**18**	27	**43**
8	25	25	14	**13**	**38**
9	25	14	2	**0**	**25**

Plotting the resulting max approach level against the calculated LFM graphically displays how the adapter card compares to the various Thermal Levels, as shown in Figure C-8.

Note that it never exceeds Thermal Level 5. Even though it is under Thermal Level 4 when approximately above approach air temperatures of approximately 55C, its MaxTherm Level would be 5. Since it could operate up through an approach air temperature of 65°C, it would use a MaxAmbient value of 65. Using a lower MaxAmbient value would not allow this hypothetical card the advantage of using a lower Thermal Level, so it should declare the maximum ambient it could support, up to the limit of 65°C.

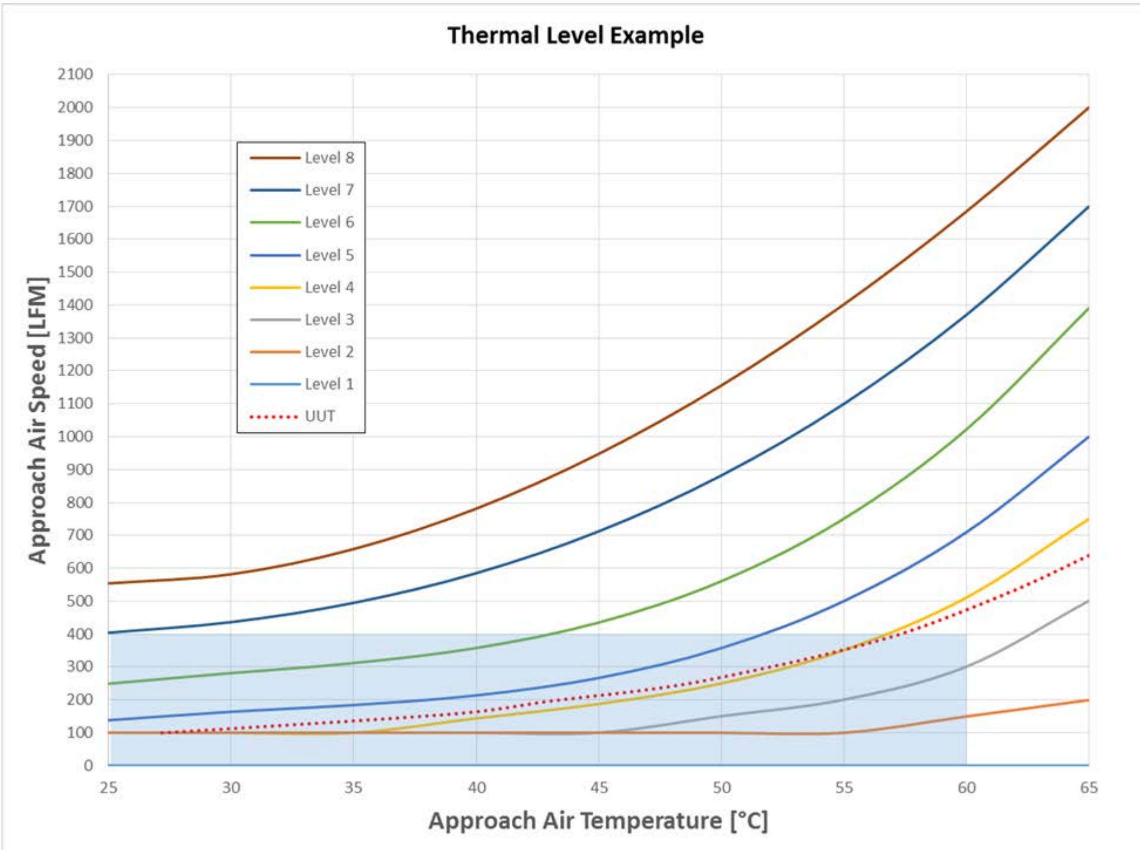

Figure C-8: Example for a Hypothetical Adapter's Thermal Level

Since this hypothetical adapter card does not support lowered power levels / lowered performance when under adverse environments, its DTherm Level will be the same as its MaxTherm Level.